Journal of Manufacturing Systems 74 (2024) 16–29

Contents lists available at ScienceDirect

Journal of Manufacturing Systems

journal homepage: www.elsevier.com/locate/jmansys

Technical paper

Spatial-temporal traceability for cyber-physical industry 4.0 systems

Zhiheng Zhao [a,b,c], Mengdi Zhang [d], Wei Wu [e], George Q. Huang [a,b,*], Lihui Wang [f]

[a] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, China
[b] Research Institute for Advanced Manufacturing, The Hong Kong Polytechnic University, Hong Kong, China
[c] State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, China
[d] Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong, China
[e] College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing, China
[f] Department of Production Engineering, KTH Royal Institute of Technology, Stockholm, Sweden

ARTICLE INFO

Keywords:
Spatial-temporal traceability
Indoor positioning
Cyber-physical industry 4.0 systems
Spatial-temporal reasoning
Cyber physical internet

ABSTRACT

The COVID-19 outbreak has posed significant challenges to end-to-end global supply chain visibility and transparency, with city lockdowns, factory shutdowns, flight cancellations, cross-border closures, and other uncertainties, disruptions, and disturbances. To address these challenges, reliable and accurate spatial-temporal information of physical objects and processes is essential to understand the industrial context and predict potential risks or bottlenecks for further decision-making. Product traverse both indoor (e.g., shopfloors and warehouses) and outdoor (during transportation) contexts. Despite significant advances in spatial-temporal traceability for outdoor environments using Global Positioning System (GPS) and Geographic Information Systems (GIS), satisfactory performance has not yet been achieved in indoor context, which accounts for the majority of operations. This limitation results in disjointed visibility and inaccessible transparency across the holistic supply chain. This research introduces universal and interoperable spatial-temporal elements for cyber-physical industrial 4.0 systems (CPIS) and develops a multi-modal bionic learning (MMBL) method for accurate and enduring indoor positioning. Proximity, mobility, and contextual reasoning mechanisms are designed to capture the interplay, evolution, and synchronization among objects at the operations level. To validate and evaluate the effectiveness of the proposed solution, we first conduct laboratory experiment and then apply the method in a real-life case company. Comparative analysis is conducted. MMBL clearly outperforms the other methods with 95% of the errors are within 3.41 m and maintains effectiveness after a year of use, which represents a significant step forward in achieving spatial-temporal traceability in CPIS.

1. Introduction

The purpose of Physical Internet (PI) is to manage the flow of physical object just as the data packets that transmitted in the Digital Internet (DI), which attracted much attention from various stakeholders including academics and practitioners [1]. It seeks to establish full interconnectivity among independent logistics networks and services through intelligent interoperability for resource and service sharing [2]. The work in [3] has reported several benefits of PI, including increased profits, lower prices, and reduced pollution, through the development of a highly interconnected global logistics system. While the DI provides commands such as the "PING" (Packet InterNet Groper) to determine reachability and "traceroute" to display possible routes and measure delays of data packets from source to destination, there is a lack of dedicated commands, standardized measuring and representation, and supportive analytics to monitor, track, and trace physical objects in the PI. The uncertainties and disturbances in interweaved indoor and outdoor global supply chain activities put a strain on the satisfactory delivery. The interrupted port operations, cancelled flights, delayed shipment arrivals and departures during COVID-19 outbreak have substantially constrained capacities and created serious operational jams and deadlocks at terminals and ports. Global supply chain logistics suffers from lacking end-to-end supply chain visibility, and information traceability and transparency [4]. The need for shared traceability to monitor performance among various parties is undoubtedly reinforced [5].

The Cyber-Physical Internet (CPI) is a term that emerged from the Hong Kong Theme-based Research Scheme Project, "SynchroHub:

* Corresponding author at: Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, China.
E-mail address: gq.huang@polyu.edu.hk (G.Q. Huang).

https://doi.org/10.1016/j.jmsy.2024.02.017
Received 26 April 2023; Received in revised form 26 January 2024; Accepted 26 February 2024
Available online 1 March 2024

Cyber-Physical Internet for Synchronizing Cross-Border Logistics Hubs in the Greater Bay Area (GBA)". The goal of CPI is to establish a system for sending and receiving goods that is like sending and receiving messages within instant chat groups, with a focus on providing great spatial-temporal traceability for all group members. Compared to the PI, which mainly concentrates on the physical logistics interconnection aspect, CPI places more emphasis on spatial-temporal analytics. This is because not only does physical traceability form the foundation for explaining the synergy of resources, but many operational optimizations now depend more on real-time data to combat uncertainties and disturbances for resilient decisions.

Nowadays, the integration of industrial IoT, digital twin, 5G networks, and big data into cyber-physical Industry 4.0 systems (CPIS) has provided a foundation for data collection and analytics. The multi-modal sensory data generated by CPIS also provides potential spatial-temporal information about objects. This information includes the physical movements of products/materials, as well as the exact time and place of information and transaction generation. It serves as the fundamental conjunction point for the three flows - logistics flow, information flow and cash flow [6,7] in supply chain traceability. The standardized representation of specific objects' real-time spatial-temporal status provides intuitive visualization and sharing compatibility among stakeholders, which is essential for monitoring and rapid intervention during emergency situations (e.g., staff rescuing in a warehouse on fire). Traceability can also examine patterns of movement or interaction over time and contribute to identifying bottlenecks such as unreasonable layout and material flow routes. Proper traceability reasoning allows resources to understand their surroundings and take proactive actions (e.g., receiving/synchronizing tasks). Furthermore, we have observed that logistics hubs/delivery stations or assembly lines must build buffer spaces during peak hours to compensate for time fluctuations and uncertainties. The geographic distance among objects and their picking sequences not only affect time but also total cost. Therefore, good coordination between spatial and temporal dimensions is conducive to achieving zero-inventory [8] or even zero-warehousing [9].

However, challenges still exist in achieving spatial-temporal traceability in CPIS. Firstly, the widely applied Global Positioning System (GPS) and Geographic Information Systems (GIS) have paved the way for spatial-temporal traceability for outdoor environments, while most supply chain activities occur indoors, such as manufacturing and warehousing operations. The lack of a universal expression, interoperation, and sharing standard hinders spatial-temporal traceability for indoor environments. Secondly, in industrial settings, the multi-path effect and signal attenuation significantly degrade positioning accuracy. Additionally, changes in the environment over time can cause accuracy to plummet, requiring laborious and time-consuming calibration each time the environment changes, making the development of an environment-specific positioning model challenging. Thirdly, spatial-temporal information provides valuable insights for optimization and decision-making processes in CPIS at the operation level. However, the lack of insights into patterns and trends that occur over time and space hinders more informed and predictive decision-making. To address these challenges, the following research questions have been formulated:

(1) What is the most appropriate representation standard to achieve shared and interoperable spatial-temporal traceability considering the shuttling of objects between indoor and outdoor environments?
(2) How can accurate and reliable indoor positioning be achieved in an enduring manner using multi-modal data from CPIS?
(3) How can spatial-temporal data be leveraged through reasoning mechanisms to support decision-making related to operations in CPIS?

In this article, we address the aforementioned challenges by first defining and comparing basic spatial-temporal elements for geospatial traceability in both indoor and outdoor settings. We then propose a multi-modal bionic learning method inspired by biological cell evolution and mutation to achieve accurate and reliable indoor positioning. Finally, we introduce three types of spatial-temporal reasoning mechanisms to generate insights and predictions that support intelligent decision-making in CPIS. To verify the effectiveness of our proposed solution, we implement the method in the shopfloor of a leading computer equipment manufacturing company.

This paper is organized as follows. Section 2 reviews spatial-temporal traceability related research in CPIS. Section 3 discusses the basic spatial-temporal elements for traceability with unified representation and measurements. Section 4 introduces the multi-modal indoor positioning method and related spatial-temporal reasoning mechanisms. The proposed solution is first tested in laboratory settings in Section 5 and then verified and evaluated in a real-life case study in Section 6. Finally, Section 7 concludes the work.

2. Literature review

The literature on the spatial-temporal traceability for CPIS is extensive, covering a range of topics from acquiring the spatial-temporal information to generating spatial-temporal related decision. In this paper, we review the existing literature on spatial-temporal traceability in terms of data expression and representation, indoor positioning in CPIS, and spatial-temporal related decision-making in CPIS applied areas.

2.1. Spatial-temporal data expression and presentation

The spatial-temporal data are data that imply space and time information. For the spatial data, the problem of getting lost in outdoor environments has been significantly reduced with the use of GPS and GIS. These technologies have revolutionized the way we navigate and explore the outdoors, allowing us to accurately locate position on a map application and plan routes accordingly. The satellite-based positioning system locate the objects that equipped with specific chips with longitude and latitude results through trilateration, a ranging method [10]. The normative geospatial data lays the foundation of today's location-based service applications such as Yelp, Uber, Google Maps, among others. Even the logistics industry gains great benefits from outdoor transportation traceability through the GPS/GIS [11] as customers can track their cargos with graph user interface. However, due to the long-wave signal degradation factors, GPS cannot work properly in indoor environments especially for shopfloor and warehouse where CPIS operates, which renders the sharing and interoperability for indoor spatial-temporal traceability non-standardized and inconsistent with outdoor system. Grids and Cartesian coordination system are regarded as commonly used method for geospatial data expression, presentation and manipulation which allows geospatial data to be organized, analysed, and visualized in a systematic way. The work in [12] label the environment with reference points and match the highest probability results to corresponding reference points. Cartesian coordination system is adopted in [13] to collect Received Signal Strength Indicator (RSSI) data and recurrent neural networks for accurate indoor localization with specific coordinates. The spatial distribution probability of tool in shopfloor is derived from the indoor positioning results using coordinates [14]. The experiment area is partitioned into grids and then use integrated algorithm to decide the coordinates of object [15]. As for the temporal part, the positioning real timeliness is of great importance for making agile decisions. The sampling intervals, transmission delay, and computation efficiency have been widely discussed and optimized in the literatures [16–18]. Time windows are dynamic attributes that affect the solution effectiveness. Spatial-temporal data always linked with specific interests. The Point of Interest (PoI) data is utilized in [19] to capture human factors, where an attention model that merges the

representations from the geographic and PoI perspectives adaptively is proposed. The work of [20] defined time-aware PoI recommendation which the temporal information is incorporated. A multimodal fingerprint-based indoor positioning systems is developed in [21] for airports, the georeferenced PoIs in the airports are clearly identified for explicit navigation.

Literature has yet to address the need for a unified spatial-temporal expression and presentation that enables consistent sharing and interoperability between indoor and outdoor settings. While longitude and latitude representation are commonly used for outdoor tracking, current literature tends to represent spatial-temporal data on a case-by-case basis, depending on the layout, structure, or operational context. Griding systems used in research also vary in size and shape, and the self-contained coordination system for indoor environments has limited interoperability with geographic coordinate systems. Furthermore, there is a lack of discussion in current research on the temporal relationships between information real-timeliness from a technical perspective and decision-making time window from an operations perspective.

2.2. Indoor positioning system for CPIS

Although outdoor tracking has been successfully implemented for decades, the research on indoor tracking is still progressing slowly. In current context, indoor scenarios such as manufacturing, cross-docking, and warehousing are widely existed in CPIS. The overwhelming amount of data generated through a plethora of IoT enabled heterogeneous devices post challenges to indoor tracking. The transformation from intricate spatial-temporal data to applicable and meaningful spatial-temporal information deserves more investigations. The position of picking staff is predicted in [22] for warehouse management through the RSSI values of Radio Frequency Identification (RFID). A feature selection-based back-propagation neural network that uses artificial immune system is proposed to learn the relationship between the RSSI values and position of picking staff. A RFID-enabled positioning system is proposed by [23] to locate the automated guided vehicles in smart factories. An IoT-enabled smart indoor parking system is proposed by [24] for monitoring the tractor and trailer of industrial hazardous chemical vehicles. The logistics sustainability can be achieved through reducing the searching and waiting times. The work of [25] develops a long short-term memory network-enabled tracking algorithm to locate the product trolleys via Bluetooth Low Energy (BLE). The spatial-temporal information of manufacturing resources is essential for factory logistics operations. A system is proposed by [26] to monitor worker movements on a construction site by collecting their raw spatial-temporal trajectory data and enriching it with the relevant semantic information. The signals from beacons are transformed to location coordinates. An IoT RFID data fusion procession algorithm is investigated by [27] based on spatial-temporal semantics. The data at same time but different space position and the data of same equipment (same space position) but different time are used for spatial fusion and time fusion respectively. The work in [28] clarified the graphical representation of spatial and temporal operators in the middleware to detect and disseminate events in the IoT. The entity, resource, and service models for the IoT domain is described by [29], global and local locations, temporal features are properties in the models for matching and integration purposes. The aerial vehicles and robots require the support of localisation and path planning to conduct operations in the manufacturing scenario [30]. A real-time localization of unmanned aerial vehicle is studied through computer vision capabilities [31]. An infrastructure free indoor positioning system is implemented in [32] for light object logistics and missing tool search of autonomous aerial robots in industrial manufacturing. 5G/visible light communications (VLC) tracking solutions in Industry 4.0 is reviewed by [33] from human-centric perspective. Compared to electromagnetic tracking solutions, the 5G/VLC prevails in line-of-sight scenarios. A novel human localization method in [34] is developed for robotized warehouse, operators wear vests equipped with visual sensors to localize the ground markers by fusing stereo visual-inertial odometry data and distance.

Existing research focuses on transforming one of certain specific signals attached to a particular resource into location results, which is often limited to the individual sectors of operation. However, in CPIS, numerous sensors are equipped for various purpose, and more in-depth research is needed to leverage existing end nodes to improve positioning accuracy. Additionally, the widespread presence of metallic objects and concrete walls in CPIS significantly degrades the positioning accuracy, and the change of layouts affect the endurance of indoor positioning system [35], posing challenges to realizing long-term spatial-temporal traceability.

2.3. Spatial-temporal reasoning and decision-making

In recent years, the academia has witnessed a surge in the number of spatial-temporal traceability in technology-driven Industry 4.0 era. The traceability does not only consider the trace and track of the object, but also contains context related reasoning for real-time situation calculation and resilient responses on product safety, quality, and operator behaviours [36] for value-driven Industry 5.0. The work of [37] discusses how location and time can support smart manufacturing, movement data and quality measurements are visualized and analysed using spatial-temporal analysis to compare behaviours. The spatial-temporal data are fed to the neural network frame for construction intelligent physical system [38]. The recurrent neural network structure is linked to predict the mobility of an object, and the 3D convolutions for modelling the relationships between objects. The work of [39] proposes event attributed spatial entity knowledge based to formalize spatial entities in a geographic region whose temporal attributes are events to answer the event-based queries on prediction of spatial process. Moreover, spatial-temporal information has also been considered for case-based reasoning in disaster management by measuring spatial-temporal similarities [40].

Data generated in cyber-physical systems are dynamic, volatile, interdependent, containing rich spatial-temporal information, and they are critically important for real-time decision making [41]. A spatial-temporal out-of-order execution which divides the space and time scopes of a factory into finite areas and intervals to reduce complexity of optimization problem [42]. A hedging coordination in prefabricated housing production is studied in [43], the double handling cost is mitigated substantially through spatial-temporal method by reducing assembly time uncertainty and on-site congestion probability. A novel spatial-temporal analytics is developed to identify the high spatial-temporal correlated but indirect contact in the COVID-19 situation [44]. The spatial and temporal asynchronization in production and logistics may cause ineffective resource allocation, production scheduling and even safety threats to human operator. A dynamic spatial-temporal knowledge graph is introduced to allocate production logistics resources where the IoT big data is analysed to generate spatial-temporal information through deep learning [45]. The timely spatial-temporal traceability and visibility with cyber-physical synchronization is actualized. An automated fault detection and isolation approach in [46] models the spatial/temporal production data with high accuracy.

Many studies of resources synchronization and cooperation exploits the static attributes of resources (type, capacity, fixed cost, availability) while neglecting dynamic spatial-temporal relationships as it is hard to obtain and quantify. Spatial-temporal information implies the movement rules and patterns of resources which can be inferred to valuable insights such as bottleneck identification and process optimization for managerial implications. Moreover, while current research on decision-making processes largely harnesses operations data, little attention has been devoted to enabling resources with contextual awareness through spatial-temporal data to self-understand the process and provide support

for making local decisions.

3. Basic spatial-temporal elements in spatial-temporal analytics

3.1. Spatial grids and temporal units

The timeline is discretized with time windows. Likewise, a space is discretised into 2D grids or 3D cubes. Discrete grid systems are critical to analysing large spatial data sets, partitioning geospatial areas into identifiable grid cells to support efficient real-time location-based quiring services. A set of basic spatial primitives (e.g., point, line, polygon) is utilized to model and abstract reality accordingly. The referencing and indexing mechanisms provide reliable methods to access, store and retrieve data. Discrete global grid systems such as Uber H3 and Google Map S2 have been validated with commercial success [47,48]. The industrial site is more complex and in a constant state of change than the urban society. However, most CPIS operations are carried out in large scale multi-storey indoor plants. Current discrete geospatial grid systems are deployed globally oriented to outdoor environments. The indoor spatial analytics require proper indexing system to identify both indoor and outdoor geolocation elements.

One of the main differences between S2 and H3 is the choice of cell shape. S2 uses square cells while H3 uses hexagonal cells. Both systems do not need to understand the industrial site context. The important differences lie in neighbour traversal, subdivision, and visualization.

Neighbour traversal: There are only three shapes that can form regular tessellations: equilateral triangles, squares, and hexagons. The pivot of the hexagon cell (HC) to other neighbours are equidistant while triangle and square cells have three or two different distances. Hexagons allows for simpler analysis of movement. The spatial temporal traceability of the industrial asset often concerns with the proximate assets to realize synchronous optimization. Compared to square raster, hexagon is 13% more efficient at sampling and 25% to 5% more efficient for common image processing algorithms [49].

Subdivision: Hexagons do not cleanly subdivide into seven finer hexagons. However, by alternating the orientation of grids a subdivision into seven cells can be approximated. Consider warehouse management as an example. Typically, the assignment of storage locations is determined based on the product category. The further division of the storage location is different product types but belonging to the same category. The inclusion relation is continuously inherited following a multilevel parent-child pattern. On the other hand, the smallest possible resolution should be indexed and is able to aggregate into larger resolutions. Even square grids can have perfect subdivision, however 7 hexagons 1/7th the area covering the exact same amount of area with a known amount of error. The shape can be approximated with 19.1 degrees of rotations. The average hexagon area is 0.9 m^2 at maximum resolution level 15 in H3, which can satisfy the location accuracy in the industry applications. Furthermore, the localization accuracy in the Industry 4.0 varies according to different operation requirements.

Visualization: Hexagons can tessellate the plane regularly. The cells can appear distorted under Web Mercator Projection. According to [50], H3 cells have the same non-alignment with the map projection, but the effect is less noticeable to viewers for hexagons. In industrial operation level, the effect can be almost neglected. Moreover, users with different privileges can access to varying resolution levels. For example, the operators in the organization may have access to higher resolution levels while the customer or partners who wish to track the products only can access to lower resolution levels since privacy concerns.

However, the H3 system ignores elevation and treats all points with the same latitude and longitude as the same location. This poses a challenge in indoor environments where multi-storey shop floors or warehouses are commonly existed in CPIS. Inspired by the addressing of computer networks in Internet where Classless Inter—Domain Routing (CIDR) is fully utilised to divide IP address into a hierarchy of subnets of assorted sizes, we blend self-contained coordinate system into the H3 griding system with extra height indicators in Fig. 1. The end devices on the Internet have IP addresses to identify network interface identification and location addressing through router assignment. The physical address, a 12-digit hexadecimal number assigned to each device, is the unique identification for every end device. In the cyber physical internet, the location of the objects is estimated through positioning method using Cartesian coordination system. We first map the coordinates derived from indoor positioning results to the H3 hexagon cells at predefined resolution levels according to accuracy requirements. The hexagon cells are like exact IP address in the computer networks where lower subnet mask in CIDR indicates larger covering area. Then, we label the hexagon cells that tessellate in the building with height indicator to show the level information. The gateway, acting as router in the computer networks, have ability to assign the positioning results of varied objects through embedded sensors and positioning engine. Each object equipped with wireless sensors carries a physical address as Media Access Control (MAC) address in datalink layer of computer networks. It is the unique identification for the object in CPIS.

Fig. 2 depicts two types of time windows. The first pertains to the intervals at which IoT devices or sensors gather real-time data and update the condition and state of the factory. We can describe this type of time intervals as heart beats τ of a smart factory as the signals/data/information generated proactively following a certain rhythm. It determines how real-time a factory is. Many researchers have concluded that there are basically three to four layers of IoT architecture from collecting raw data to generate meaningful information in Industry 4.0 [51,52]. We take the location estimation as an example; the location of the moving objects is determined by the calculation results based on the received signal strength indicator (RSSI). There are also three parts of IoT supportive hardware at cloud, edge, and end side, respectively. The IoT tag that attached to specific object at end side emits signals in a certain frequency τ^b. Multiple RSSI readings collected by the gateway at edge side through constant sniffing are reported to centralized server in certain interval τ^g. The centralized server at cloud side receives the data from gateways and analyse to update the global moving object's location status in interval τ^s. As the signals are prerequisites of the location estimation, even the interval of heart beats is adjustable at each side, but the interval must compile the following constraints: $\tau^b \leq \tau^g \leq \tau^s$.

The other type includes the time intervals at which the smart factory makes and updates its decisions following planning $\mathscr{T}$, scheduling T and execution t regime. We describe this as decision time window of a smart factory. If $t_1 = \tau_1$, this is the case where factory updates its real-time status and make its real-time decision at the same pace during execution level. If $t_1 = n \times \tau_1$ where n is integer, factory updates its status n times before making a decision. If the factory's status is not updated (changed), there is no need to make another round of decision. A decision cycle is only triggered by changes of status due to one or more beats. We borrow the notation of CPU (Central Processing Unit) as the "brain" of CPIS. Likewise, CPU clock is introduced to measure the speed or frequency at which decisions are made. The CPU will assign and execute ready operations to suitable time periods which include at least one- or multiple-time windows. The planning interval $\mathscr{T}$ usually considers longer period of time including setting goals and determining the overall strategy for achieving the goals, while the scheduling interval T is much shorter than planning interval. The scheduling usually considers the allocation of resources, machine utilization, and production line sequencing. The execution interval t is relatively small and very much related to real-time pointer for agile decision-making. Some computer manufacturers adopt fixed execution interval such as 2 h to conduct material preparation before production. This cycle repeats as the clock ticks just like the CPU clock.

3.2. Spatial-temporal Points of Services (POS)

A point of interest (PoI) is a specific point location that someone may

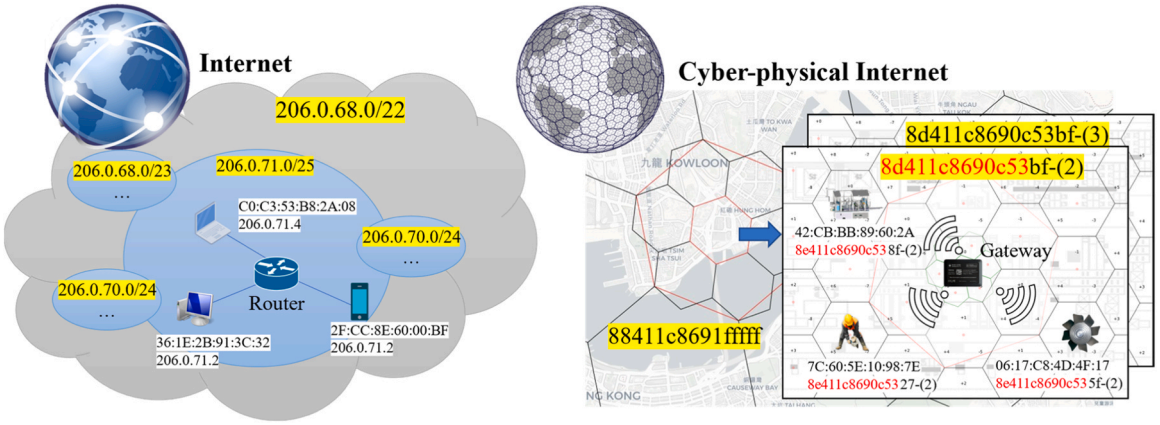

Fig. 1. Addressing in Internet and Cyber-physical Internet.

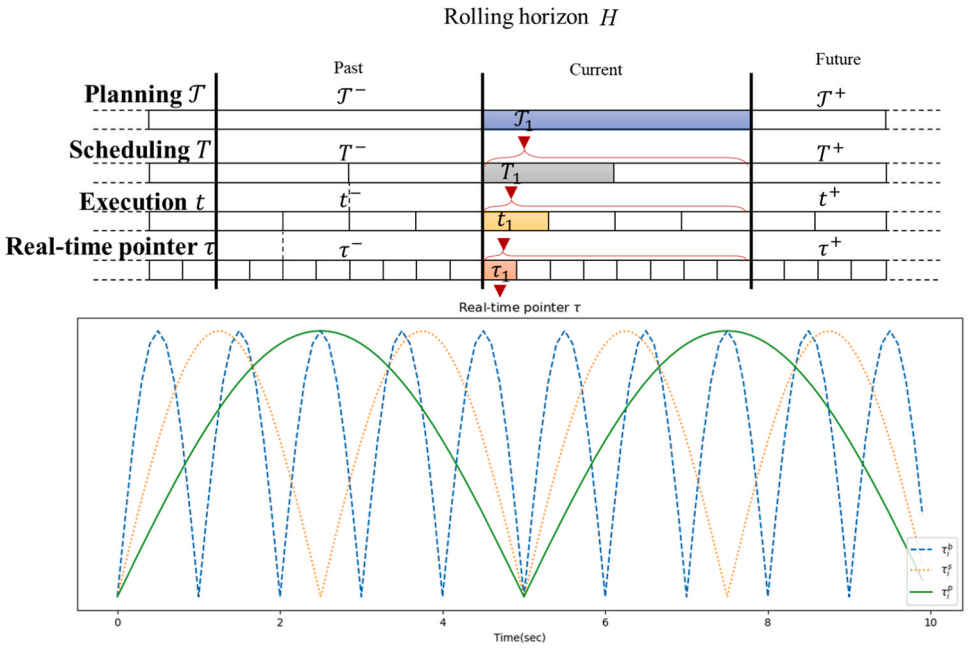

Fig. 2. Factory clock in CPIS.

find useful or interesting and can be shared through location-based social networks such as Foursquare, Facebook Places, and dianping.com [53]. PoI data represent various venues in human society including shopping malls, coffee shop and parks and widely adopted in various web mapping platform especially location-based recommendation applications. The CPIS also concerns vast number of points where events and transactions conducted. In analogy to visiting the museum at its opening time, the trucks are only allowed to unload the freight during working hours of receivers. Earliness or tardiness not only affect the successful completion of the task, but also incur extra labour cost and expenditures. The spatial-temporal window of the PoS significantly affect internal and external operations planning, scheduling, and execution. We define the spatial-temporal point of service (PoS) is the place where industrial events or transactions conducted with time-awareness.

PoS are dynamic. First, PoS of a spatial element change with time. For example, the machine located in a spatial element operates on component A at a time and on component B at another time. This is the norm in manufacturing systems. Second, some factory objects are mobile rather than stationary. PoS of a mobile object may change from one location to another. For example, PoS of a trolley is to 'load' work-in-progress (WIP) items at the source machine centre and to 'unload' WIP items at the destination machine centre, and to "carry and move" on the path in between the two locations. Third, PoS of smart objects in CPIS may be mutually changed when they interact and/or interoperate with each other. For instance, human operators may conduct different tasks at different locations and times. Assuming the materials in workstation (PoS) are waited to be packed, the human operator proceeds to workstation first, and provide packaging services there afterwards, the human operator become PoS for subsequent materials to be packed in the future.

Inspired by the digital twin conception [54], we model the PoS and define it as $M_{PoS} = (ST, PE, VE, CN, Ss)$. The model of PoS consists of five parts. The ST refers to the spatial-temporal information including current location, service starting and ending time, and historical data for storing, querying, and sharing. PE defines the physical entity which includes the functional information $FI = (ID, name, FunctionDescription, PiC)$ such as ID, name, functional description, and person in charge, etc., and resource-associated information $RI = (ID, Type)$ as some of the point of service may associated with certain resources. The functional information is usually static information, but the resources-associated information may be dynamic. For example, a storage location (without RI)

is the PoS for keeping materials, a static CNC machine (with static *RI*) is the PoS for processing, and movable forklifts (with dynamic *RI*) are PoS for transporting materials. The *VE* part concentrates on the virtual entity of geometry visibility attributes such as shape, size, and display related behaviours. The *CN* specifies the connections between the physical and virtual entities through various IoT technologies and necessary network communications. The services *Ss* is highly correlated with the context-based operations. It analyses the data collected to infer implicit information such as real-time status of resources (idle, occupied or breakdown), level of proximity between the user to the PoS, and estimated time or routes, etc. The services also contribute to the credible traceability sharing among different parties as the PoS plays the role of eyewitness during the spatial-temporal reasoning.

4. Spatial-temporal traceability through multi-modal bionic learning in CPIS

The core of spatial-temporal traceability is to identify the location and time of existence of an object. Cyber physical Industry 4.0 applications involve both indoor and outdoor operations. Global Navigation Satellite System (GNSS) such as Global Positioning System (GPS) and BeiDou Navigation Satellite System (BDS) can be readily used for outdoor wide-ranging positioning. In CPIS, operations are mostly conducted indoors, especially for manufacturing and warehouse operations. For the majority of operations that occur indoors, a suitable indoor positioning system is essential. In this section, we first discuss the spatial-temporal traceability enabled by a multi-modal bionic learning-based indoor positioning system, which facilitates the precise and reliable localization of objects. The overall process of multi-modal bionic learning for indoor positioning system, which is analogous to the process of cell mutation, crossover, and evolution in biology is introduced. We then delve into the specifics of the multi-modal bionic learning (MMBL) algorithm used in IPS. We then propose mechanisms of spatial-temporal reasoning in the last of this section.

4.1. Analogical process of multi-modal bionic learning for indoor positioning system

In CPIS, different shop floors and warehouses in industrial parks may have different information communication technology infrastructures and shareholders. For example, the transportation among supply chain players usually involves outdoor transportation where GPS/BDS have certain merits and convenience. For the warehouse management, the RFID reading events, considered as reactive positioning, records the "process location" instead of absolute spatial-temporal information. BLE and Wi-Fi share the similar tracking principles as the access points deployed at fixed locations to capture the signals and send back to the server. The GNSS (GPS/BDS) and cellular positioning (5G/LTE) already have mature positioning mechanism with absolute coordinates but still imperative for feeding the model since the transformation of absolute coordinates to hexagon cells for unified analytics, compensation for single signal reasoning, and time synchronization. In this research, we integrate all the above-mentioned technologies which have potentials for directly or indirectly realizing spatial awareness and combine them as radio frequency feature collections (RFFC) for inferring the matched hexagon cell.

Fig. 3 showcases the overall process of MMBL in the light of biological cell evolution. The left side of this Fig. tells the story of how the biology cell is cultivated, extracted, judged, and evolved, while the right side depicts the corresponding process implemented in the hexagon cell localization. MMBL consists of two stages, which are offline calibration stage and online prediction stage.

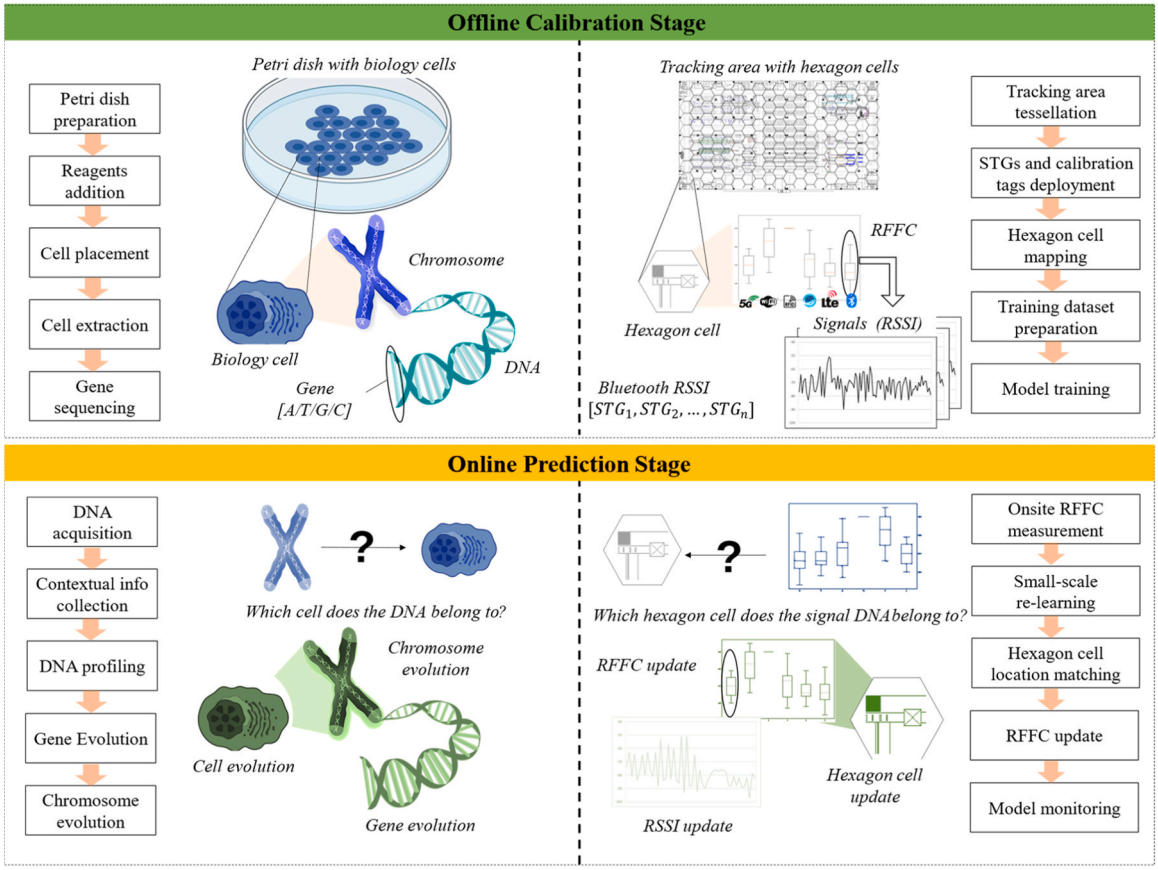

Fig. 3. A biological analogy to MMBL process.

In the offline calibration stage, with the aim of correct identifying the type of each biology cell (muscle, epithelial, adipocytes, etc.), the features of chromosome, DNA, and genes of the cells are supposed to be clearly extracted and labelled in the first place. The petri dish that cultures the observes diverse cells is prepared with adding reagents. Varied types of cells are placed in the petri dish for further extraction and gene sequencing. Chromosomes are structures in cells that contain genetic information in the form of DNA, along with associated proteins and other chemical substances. Together, these components make up the fundamental building blocks of biological cells. DNA is the molecule that carries the genetic information that is passed down from one generation to the next, while genes are specific segments of DNA that code for specific traits or functions. The specific sequence of nitrogenous bases in a gene, which are adenine (A), thymine (T), guanine (G), and cytosine (C), provides unique and distinctive information that distinguishes one gene from another. For the hexagon cell positioning problem, the tracking area is first confirmed and tessellated with hexagon cells. A spatial-temporal gateway (STG) is a multi-source signal receiver with additional edge-computing capacity for analysing and categorizing initial readings. STGs are deployed in regular intervals in the tracking area, without the need for intentional coordinates marking. In this case, implementation is less laborious. Calibration tags are placed at the centre of each hexagonal cell to broadcast multi-modal specific radio frequency signals, such as Bluetooth Low Energy (BLE) or Wi-Fi, which can be received by the STGs. The mapping from the hexagonal cell to Cartesian coordinates is relatively simple and can be inferred from the regular graph tessellation. The STGs collect radio frequency readings during certain time windows and record them as dataset for model training. Some of the calibration tags which can be easily installed without affecting daily operations can be permanently deployed at fixed locations and redefined as reference tag for online prediction purpose. It is noteworthy that the RFFC has more than one source radio frequency features, and each type of the radio features is represented as signal DNA with a set of received signal strength indicator (RSSI) readings from all STGs to dedicated hexagon cell. Therefore, a segment of RSSI from single STG is formed as one single piece of gene. The RSSI is very jumpy in the industrial settings as the metallic objects and concrete walls cause multi-path and none-line-of-sight (NLOS) issues. Several signal filtering techniques such as Kalman filters and particle filters are employed to smooth the fluctuating signals.

During the online prediction stage, as analogous to identifying criminal suspect from DNA samples and contextual information in criminal investigation, we employ the pre-trained model to derive the location estimation through the online measurements of RFFC while simultaneously considering readings from reference tags which reflects onsite environmental factors. The changing layouts may significantly affect the environment-specific model performance. Therefore, we adopt small scale learning to keep the model understand the newest environment before location estimation. We observe the real-time RFFC from reference tags and make modifications to update the initial training set before the location prediction. The deep neural network-based model conducts small-scale re-learning to make weights and bias more approaching the on-site environment. After prediction, genes in DNAs may have mild change (evolve) or distinct change (mutate) according to the observation which brought by the environmental influences. The RFFC are also updated after each time of positioning with both correct labelled data and unlabelled data for further feature extraction [55].

4.2. Mathematical model of multi-modal bionic learning for indoor positioning system

We consider a tracking area A with N_g deployed spatial-temporal gateways (STGs) and Q hexagon cells tessellated in. Spatial-temporal gateways are a type of IoT device that is specifically designed to capture and process spatial and temporal information from a variety of sensors. These gateways can receive signals from add-on sensors such as Bluetooth Low Energy (BLE), Wi-Fi, RFID, GPS modules, and other similar sensors. With their advanced processing capabilities, they are able to reason about the data they receive, identify patterns and make decisions based on the insights gained. We define $N_g = N_b + N_w + N_r + N_o$ where N_b, N_w, N_r, N_o are the number of BLE, Wi-Fi, RIFD, and other sensors respectively. We exemplify the BLE signal calibration in the following details. Total N_{ct} calibration tags are temporally deployed in the centre of hexagon cells to fulfil calibration task before online positioning. Each hexagon cell has a mapping relationship with Cartesian coordinates in two-dimensional environment. The position of calibration tags is defined as $p_q^c = \left(x_q^{'}, y_q^{'}\right), q \in (1, 2, \ldots, Q)$. It is noteworthy that coordinates of the equally divided partitioning on regular graph is easy to be reasoned. The calibration tags emit different types of radio signals to nearby STGs. Each STGs collects the RSSI measurements broadcasted by the calibration tags.

The rolling horizon H is divided into equal time windows $\tau_t, t = 1, 2\ldots, T$ as real-timeliness pointer, one round of positioning is required to be finished within in τ_t. Let $\Gamma_{n_b}^{\tau_t} = \left(r_{1,n_b}^{\tau_t}, r_{2,n_b}^{\tau_t}, \ldots, r_{Q,n_b}^{\tau_t}\right)$ be the row vector of BLE RSSI received by STG n_b from each hexagon cell from 1 to Q at time window τ_t, where $r_{n^b}^{\tau_t}(n_{ct})$ is the filtered BLE RSSI value measured by STG n_b from calibration tag n_{ct} at time window τ_t. Multiple rounds of calibration signals are required to collect to ensure the sufficient training data. Unavailable signals are filled with very small value. For passive reading actions such as RFID, successful reading events are typically indicated by large RSSI values. With collection of various signals such as Wi-Fi Γ_{n^w}, RFID Γ_{n^r} or other sources Γ_{n^o} after multiple rounds, we transform training data set to $X_{train} = \begin{bmatrix} | & | & | \\ x^{(1)} & \cdots & x^{(m)} \\ | & | & | \end{bmatrix}$, $X_{train} \in \mathbb{R}^{N_g \times M}$ from perspective of hexagon cell where $x^{(m)}$ defines mth training data from all STGs. The corresponding $Y_{train} = \begin{bmatrix} | & | & | \\ y^{(1)} & \cdots & y^{(m)} \\ | & | & | \end{bmatrix}$, $Y_{train} \in \mathbb{R}^{2 \times M}$ is the coordinates of the hexagon cell where the training data is collected and classified. To evaluate the model's performance, we use k-fold cross validation in conjunction with a hold-out validation set. For each fold, we reserve 20% of the data as a separate validation subset to assess the model's generalization ability. Many traditional ranging algorithms employs the log-distance path loss model to derive the distance according to the RSSI readings, however, in the industrial settings, different environment may significantly affect the path loss exponent rendering the estimation unreliable.

The regression function of deep neural network is adopted to estimate the (x, y) coordinates of the unknown targets. We define the action value of lth layer of the model as $A^{[l]} = \sigma(W^{[l]}A^{[l-1]} + b^{[l]})$, where $W^{[l]}$ and $b^{[l]}$ are the weight parameters and bias value of layer l. The activation function σ includes *sigmoid* $f(x) = \frac{1}{(1+e^{-x})}$ and *ReLU* $f(x) = \max(0, x)$ is employed to introduce no-linearity into the neural network, which allows it to learn and model complex relationships between input and output data. The output layer has only 2 neurons which represent the x and y values of the final location estimation results.

The loss function between ground truth coordinates (x, y) and predict value $(\hat{x}, \hat{y})$ is set as Mean Squared Error $\mathscr{L} = \frac{1}{M} \times \sum_{i \in M}((x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2)$, where M is the total number of training data in the dataset. The MSE is a commonly used loss function in regression analysis that is particularly relevant in the context of indoor positioning. MSE is preferred over other loss functions because it penalizes large errors more heavily than small errors, which is in line with the error tolerance requirements in the indoor positioning process. Specifically in indoor positioning, a system is typically more tolerable of slight errors in location estimation, such as mispositioning to adjacent hexagon cells, as

compared to significantly distant estimations. Additionally, the MSE is differentiable, which makes it useful for gradient-based optimization algorithms.

To minimize the loss values at each training epoch, gradient descendent is leveraged through backward propagation. The weights and bias of each layer is updated through $W^{[l]'} = W^{[l]} - \gamma \bullet \frac{\partial \mathscr{L}}{\partial W^{[l]}}$ and $b^{[l]'} = b^{[l]} - \gamma \bullet \frac{\partial \mathscr{L}}{\partial b^{[l]}}$ where γ is the learning rate. We use Adam (Adaptive Moment Estimation) to compute adaptive learning rates for each parameter in the model instead of fixed ones with slow convergence during optimization process. The algorithm calculates the first moment estimate of the gradient, which is the average of the gradient values, and the second moment estimate, which is the average of the square of the gradient values. These estimates are used to update the learning rates for each parameter. The adaptive learning rates help to accelerate the optimization process.

Once the model has been trained to convergence, ensuring that there is no overfitting or underfitting issues, it can be deemed well-optimized. At this point, it is ready to make predictions during the online stage. However, accurately determining the location in industrial settings can be very challenging due to the variability of industrial layouts. This variability can lead to a significant degradation in location accuracy. Therefore, we retain some of the calibration tags to be reference tags which are easy to be placed and does not interfere with daily operations. The signals broadcasted from the reference tags each time during online stage reflects the onsite radio environment objectively. Hence, to enable the model to quickly adapt to the on-site environment, a model learning with limited updated training dataset before the online prediction is imperative. Let $X_{ref}^{t^-}$ be the radio signals collected from the reference tags before the online positioning timestamp t which we set as t^- and $Y_{ref}^{t^-}$ be the corresponding coordinates set of reference tags. We compare the labelled dataset in both $X_{ref}^{t^-}$ and X_{train} and distil the decay rate D before updating all training set X_{train} in Eqs. (1) and (2).

$$D = \frac{1}{|N_{ref}|} \sum_{i \in N_{ref}} \left(\frac{X_{ref}^{t^-(i)} - X_{train}^{(i)}}{X_{train}^{(i)}} \right) \tag{1}$$

$$X_{train}^{t^-} = \beta(D \bullet X_{train}) + (1-\beta) \bullet X_{train}, \beta \in (0,1) \tag{2}$$

where $|N_{ref}|$ denotes the total number of signal dataset from reference tag and β is the smoothing factor which typically determines the trade-off between the exploration of new values and the exploitation of previous knowledge. The training data set $X_{train}^{t^-}$ is therefore adjusted through β. A lightweighted model training is executed with the $X_{train}^{t^-}$ and Y_{train} as input with limited learning time consumption. The trained model for online prediction is therefore more adaptable to the onsite environment.

4.3. Mechanisms for spatial-temporal reasoning in CPIS

To better quantify the spatial-temporal traceability after obtaining the precise location estimation, we propose spatial-temporal reasoning mechanisms to understand how the objects interplays, evolves, and synchronize in the CPIS for improved performance monitoring, uncertainty/disturbance prediction, and resilient decision-making processes. Spatial-temporal reasoning refers to the cognitive ability to draw conclusions based on the spatial and temporal relationships between different objects or events. This involves understanding and interpreting patterns, sequences, and dependencies that occur over time and space. In a broader context, spatial-temporal reasoning can be used to predict future events, plan and optimize systems, and solve complex problems that involve elements distributed in time and space. In this subsection, proximity reasoning, mobility reasoning, and contextual reasoning under spatial-temporal traceability are illustrated.

4.3.1. Proximity reasoning

Proximity reasoning serves as a fundamental aspect of spatial analysis. It evaluates which objects are near each other, the degree of their closeness, the duration of their proximity, and the specific location where this proximity occurs. It can be leveraged to clarify the relationship between resources. Proximity reasoning is important for almost all types of internal logistics operations ensuring that the right resources are available at the right places in the right time window for the right intended context [56]. A basic formation of proximity problem is to identify objects that are within a proximate area. In a factory consisting of fixed-position assembly islands, we can apply proximity reasoning to synchronize assembly operations. An assembly operation can only be started if and only if all the necessary materials, machines and human operators are around the assembly island. To facilitate the efficiency of spatial queries and avoid ambiguous adjacency in Cartesian coordinate system. We leverage internal discrete hexagon planar grid system. It has 3 coordinate axes i, j and k spaced 120° apart in Fig. 4 where $(0,0,0)$ is its origin. The internal spatial indexing of the hexagon cells has mapping relationship with external Cartesian coordinates indexing. We define the level of proximity (LoP) between two objects u and v at timestamp t in the following equation:

$$LoP^t(u,v) = \frac{1}{2}\left(\left| {}_i^R H^t(u) - {}_i^R H^t(v) \right| + \left| {}_j^R H^t(u) - {}_j^R H^t(v) \right| + \left| {}_k^R H^t(u) - {}_k^R H^t(v) \right| \right), LoP \in N^* \tag{3}$$

where the ${}_i^R H^t(u)$ refers to the internal spatial indexing i coordinate of hexagon cell H in resolution level R that u located at time window t. The result of the level of proximity is positive integer. The LoP between the operator u in $(9,-5,-4)$ and flat trolley v in$(11,-3,-8)$ is 4 referring to the Eq. (3). In the fixed-position assembly island scenario, only if the LoP between assembly island and all the necessary materials, machines and human operators lowers than certain level the operation can be started. Fig. 5 shows that the production logistics of all required resources are completed and located within the adjacent HCs of fixed-position assembly island located in $(5,0,-5)$, in this case, the operation start command can be executed. Different applications can be derived from the LoP between different type of resources.

4.3.2. Mobility reasoning

The mobility constitutes the spatial-temporal traceability as it implies the action strategy of the objects for future analytics of layout optimization, bottleneck identification, and resource allocation improvement. The mobility has a narrow definition that describes the path of a mobile object in a given time window recording its locations and their points of services, regardless of what other objects it encounters on the path. In Fig. 4, the space-time path of forklift from $(-2,2,0)$ to $(4,-1,-3)$ is clearly marked with arrows in given time window. The direction of the arrow is determined by the time sequencies. The other form of mobility is a series of objects that go through a spatial element in given time window. Fig. 5 displays a spatial heat map of forklift mobility reasoning. Spatial elements with darker colours indicate that longer retention time of the objects. The resources' retention time-the time it stays in the spatial units reflects the traffic routes, congestion points, or operational bottlenecks. Spatial heat map in different time window also reveals the operational behaviour of the resources in workshop which contributes to the layout optimization. The mobility reasoning visualizes a step-by-step process and how objects evolved spatial-temporally.

4.3.3. Contextual reasoning

Context refers to the surroundings, circumstances, environment, background, or settings that determine, specify, or clarify the meaning of an event or other occurrence. In CPIS, context endows spatial-temporal reasoning with operational purpose and meaning. In this research, PoS of a parental location are defined as the context of the PoS

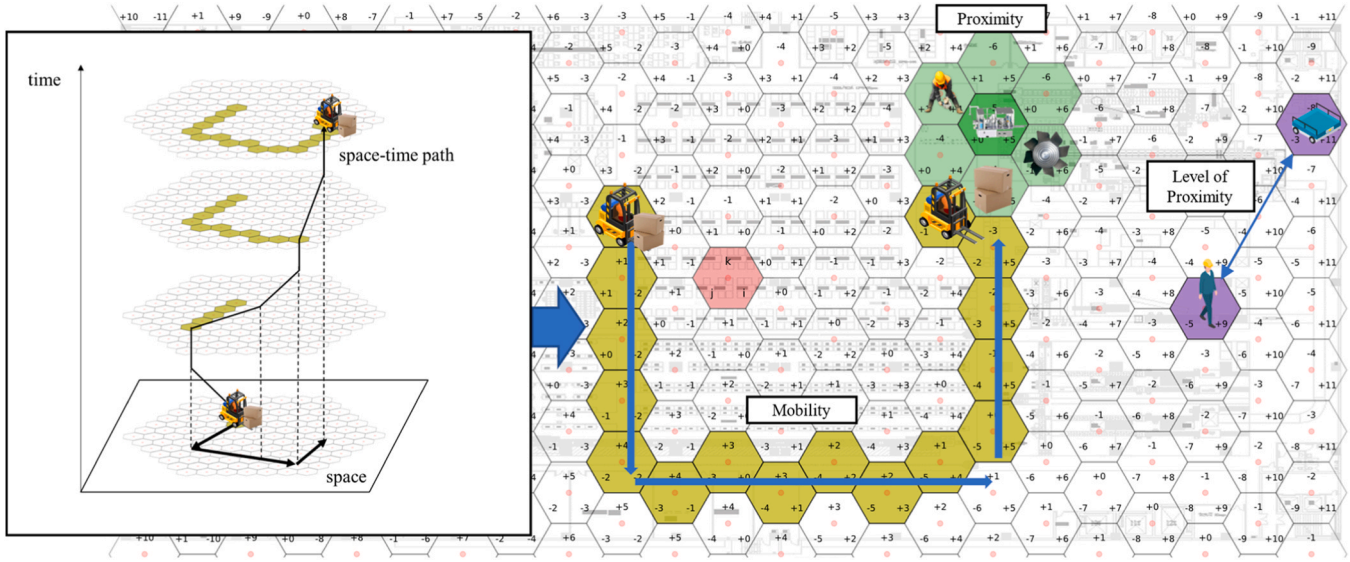

Fig. 4. Proximity reasoning and space-time path mapping.

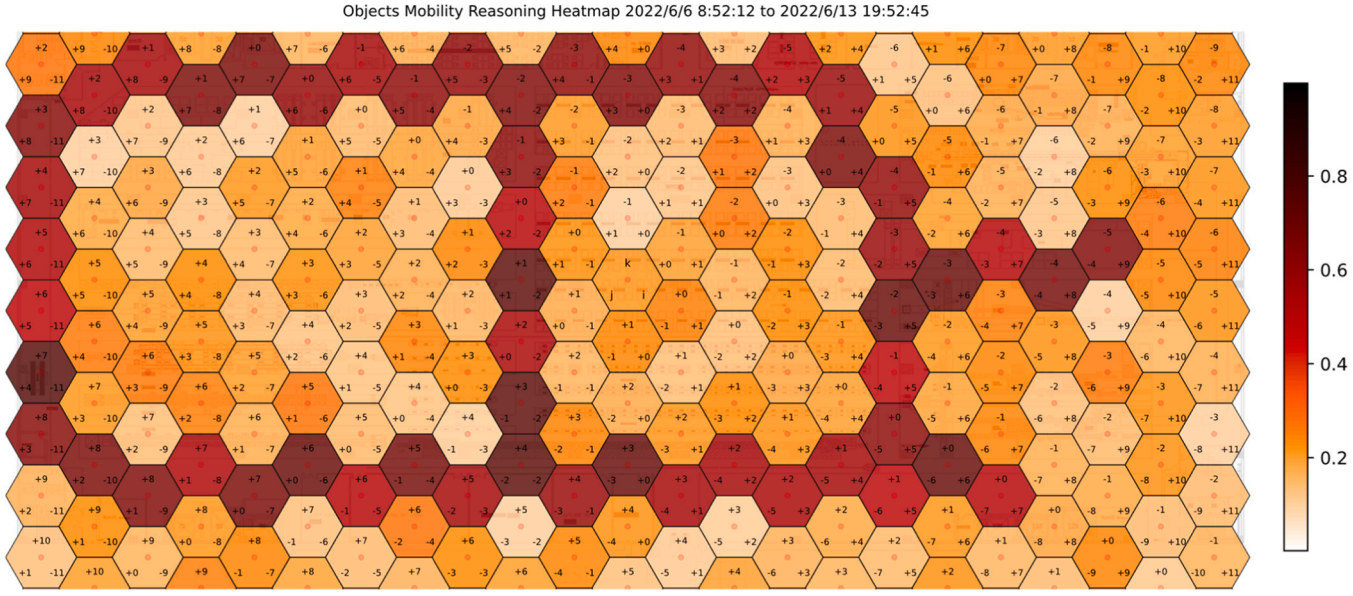

Fig. 5. Spatial heat map of objects mobility reasoning.

of its child locations. For example, the PoS of an assembly island at a time is "assembly of the right wing of the airplane" which forms the context for all the materials, machines and human operators involved in this assembly operation. These objects fall into smaller spatial grids with more specific PoS. The parent-child location hierarchies can be converted through the aggregation and subdivision of hexagon cells.

Table 1 demonstrates the possible scenarios for contextual reasoning of spatial-temporal traceability in discrete manufacturing. It concerns the various types of resources by the nature of mobility. The location-fixed resource can be regarded as PoS in parent resolution level for other resources to visit. The reasoning of a moving machine (e.g. forklift) to WIP materials located in the start/end of production line may

Table 1
Possible scenarios for contextual reasoning of spatial-temporal traceability in discrete manufacturing.

	Fixed PoS				Mobile PoS		
	Man	Machine	Material	Site	Man	Machine	Material
Man	Shifting of duty	Setup/Operation/ Maintenance	QC/Stocktaking	Homing/ Production logistics operation/ Geofencing	Contact tracing	Collaborative Operation/Task Allocation	
Machine	Task handover	Material handling	Pick-up		Safety monitoring		
Material	Replenishment/ Task handover	Production preparation	Replenishment/ Inbound		Enterprise property security		

infer the process monitoring of replenishment/pick-up operation. If both sides are moving resources, for example, a moving forklift carrying the materials, the timing of LoP change can be the action of loading and unloading. Without context, the mere spatial-temporal reasoning among objects cannot identity anomalies or draw any conclusions.

5. Experimental study

A laboratory experiment was conducted to evaluate the performance of the proposed MMBL method. The setup of the system and the data collection process are first described. This is followed by an illustration of the model construction and training process. Finally, the evaluation of the method and analysis of the results are presented.

5.1. System setup and data collection procedure

The testbed for the experiment was installed in the Cyber-Physical Internet Laboratory of the Department of Industrial & Systems Engineering at The Hong Kong Polytechnic University. The laboratory spans approximately 80 square meters ($8m \times 10m$). It serves as a workspace and study area for 10 graduate students and 21 undergraduate students. Fig. 6(a) presents the lab's layout, where four Raspberry Pi-enabled spatial-temporal gateways were set up for the experiment. Each of the STGs was mounted on a 2-meter-high telescopic tripod. For calibration purposes, roughly 120 hexagonal cells were created; at the centre of each cell, calibration tags mounted on half-meter tripods emitted BLE and Wi-Fi signals. These signals were received by the STGs, which were connected to a cloud server for data collection. Out of these tags, four were designated as reference tags to monitor signal characteristics in case of any changes in the layout. To emulate changes in an industrial setting's layout, we modified the laboratory by adding several partition panels and test where the students seated randomly and used workstations. The partition panels influenced the reflection and refraction of the signals transmitted from the tags, and the presence of the human body also caused signal attenuation. Two modes of signals including BLE and Wi-Fi are collected respectively through the STGs. There is total 62,245 pieces of RSSI data from 120 hexagonal cells collected for model training.

5.2. Model construction and training

All baseline models and MMBL were executed on a server equipped with an NVIDIA GeForce RTX 4080 graphics card. To prevent overfitting, a dropout layer with a dropout probability of 0.2 was employed. The batch size was set to 1000 for all methods, and the data were trained for a maximum of 1000 epochs, incorporating an early stopping mechanism to prevent overfitting. We partitioned the dataset, allocating 80% for training and 20% for 10-fold cross-validation. The MMBL model comprises five layers, specifically: one input layer, four hidden layers with 100, 500, 500, and 400 neurons each, and one output layer. The input layer is a 4-dimensional multi-modal RFFC as the collected data at each hexagon cell contains signals from 4 STGs, while the output layer outputs the estimated x, y coordinates of the location. The *sigmoid* activation function is applied in the first layer, with *ReLU* activation for the subsequent layers. The model employs Mean Squared Error (MSE) as the loss function, which calculates the mean Euclidean distance between the estimated and the actual ground truth coordinates. MSE is beneficial as it penalizes larger errors more significantly, aligning with the tolerance levels for location estimation inaccuracies.

Machine learning methods typically outperform traditional geometric approaches because they are more adept at capturing the correlations between RSSI data and physical locations. Additionally, historical data can provide insights into the environmental loss characteristics of the onsite situation [57]. To evaluate the performance of MMBL, we compared it with two different types of machine learning: artificial neural network(ANN) based methods, including Convolutional Neural Network (CNN) [58], Long Short Term Memory (LSTM) [13], as well as ensemble learning methods, including CatBoost [59] and

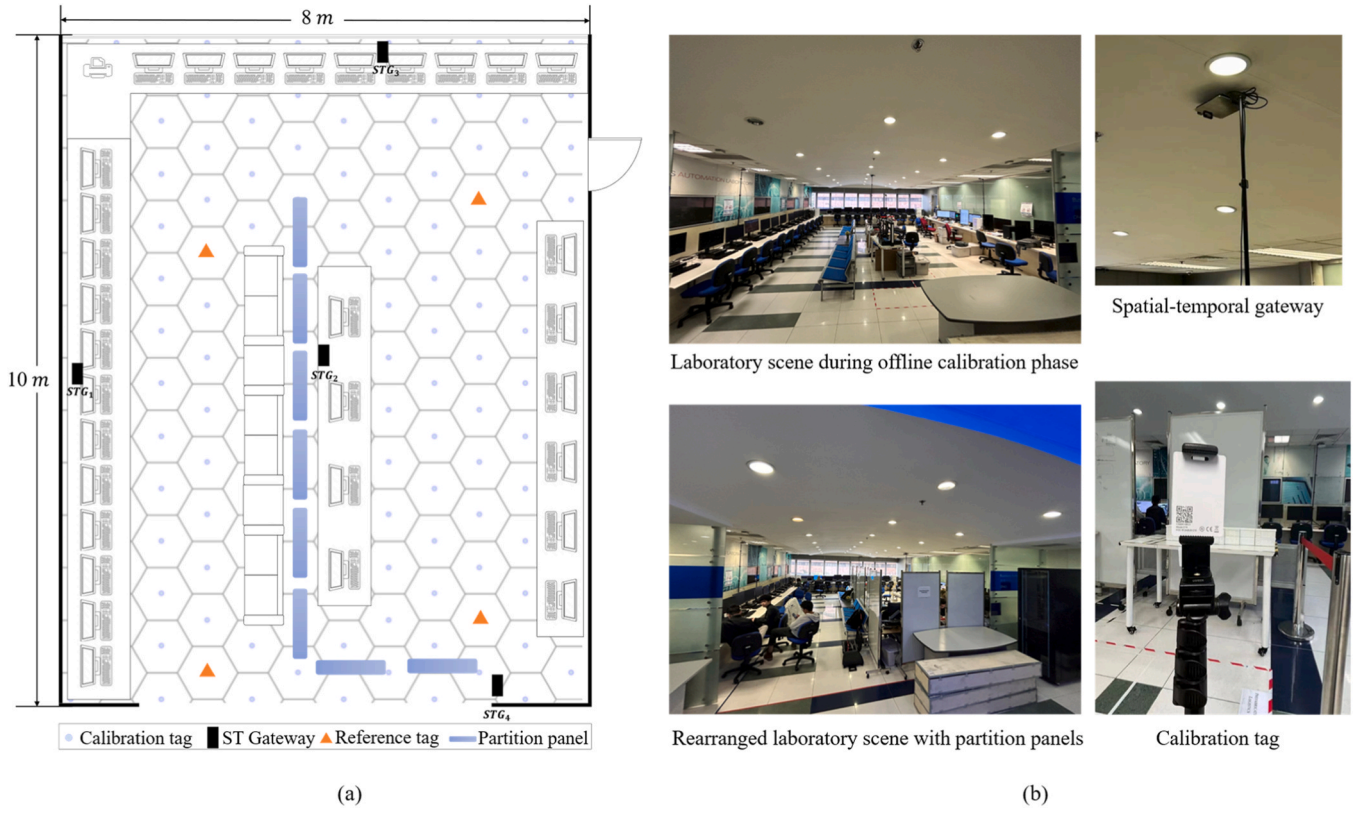

Fig. 6. Layout and implementation scene of the laboratory testbed.

AdaBoost [60]. Same datasets are utilized for the model training.

5.3. Laboratory testbed experiment results and evaluation

Fig. 7 illustrates the comparison of location errors among different methods using a cumulative distribution function from the laboratory experiment. In the figure, MMBL demonstrates superior performance, with 95% of the location errors falling within 1.44 m. This is a marked improvement over CNN, which has 95% of errors within 3.91 m, and over LSTM, which has 95% of errors within 1.32 m. Furthermore, 99% of MMBL's location errors are contained within 3.05 m in the laboratory setting. Given the reduced signal interference in the laboratory compared to the more challenging real-life Industry 4.0 environments, which are replete with metallic equipment and moving objects, MMBL's location accuracy is considered satisfactory relative to other machine learning methods.

Table 2 summarizes the positioning results obtained from different methods immediately after training the proposed model in a laboratory setting and after rearranging the layout with partition panels. This comparison underscores the robustness of the MMBL model. The data reveal that MMBL achieved the lowest mean error (ME) of 0.44 m immediately after training, outperforming CNN (1.98m), LSTM (1.07m), CatBoost (2.70m), and AdaBoost (1.96m). Even after the laboratory layout was rearranged, MMBL maintained an acceptable accuracy level, with a mean error increasing only slightly to 0.76 m. This degradation of 0.32 m in ME was the smallest observed when compared to other baseline methods. Although the LSTM method initially had the closest accuracy to MMBL, it experienced a significant decline in accuracy, increasing from 1.07 m to 1.68 m without small-scale relearning before each prediction. Additionally, the standard deviation (SD) for MMBL was 1.10 m, lower than that of the other methods, indicating a more consistent location estimation. Overall, the results demonstrate that all the artificial neural network methods (MMBL, CNN, and LSTM) outperformed the ensemble learning methods (CatBoost and AdaBoost). Further exploration of the model robustness for these methods is detailed in Fig. 8.

Fig. 8 displays the results of a 10-fold cross-validation for the five methods using a boxplot, which was conducted to assess the stability of the models' performance. Across 10 training iterations—each using a distinct subset of the training data—no outliers were observed for these methods. Notably, the median value of the LSTM model matched that of the MMBL, benefiting from the time-series information contained in the RSSI data collection. Although the ensemble learning method proved to be robust to variations in training, it did not yield satisfactory results in terms of location error.

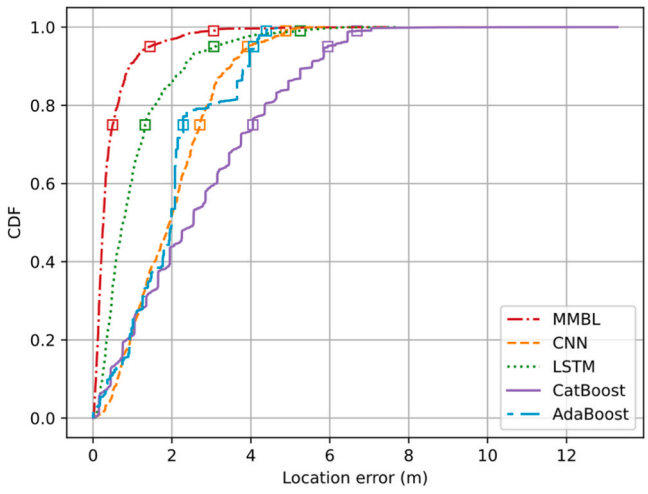

Fig. 7. Comparison of cumulative distribution function of location error for different methods in laboratory experiment.

6. Case study

To evaluate the effectiveness of the proposed methods for spatial-temporal traceability, the research team implemented the spatial-temporal traceability within a computer equipment manufacturing company located in the Greater Bay Area of China. An assembly line was selected as the test area, which was divided into hexagonal cells for the purpose of this study. STGs and calibration tags were employed for the purpose of offline calibration. The localization focus was on trolleys used for transporting WIP materials. Finally, the spatial-temporal traceability results is analysed.

6.1. Implementation scenario

The assembly line area in the computer manufacturing factory measures 56.0 m in length and 29.5 m in width, which results in a total floor area of approximately 1652 square meters. Operators assemble computer components on the main board from upstream production unit. RFID readers were installed at the starting point of the assembly line to record the accurate entry time. The assembled computer products will be stacked on the shelves of trolley. The products on trolleys will be delivered to designated testing area only if the testing order received. Hence, apart from transportation, the trolley is also regarded as buffer for temporary stowing of computer assets since no storage room are allowed. However, the trolleys are highly mobile which pose challenges for localization if transportation order arrives. Logistics operators must find the specific trolley, which is laborious and time-consuming. Coupled with that, the situation of assembly line either congested or unbalanced which cannot be clearly obtained and monitored.

Fig. 9 describes the factory layout and deployment plan. We mapped 190 hexagon cells (19×10) to cover the assembly line area with around 2.95 m as its inner diameter. 3 RFID readers were originally deployed for recording the entry of the main board inside of a tray. 12 STGs are mounted from the ceiling using a telescoping pole where power over Ethernet is provided through. BLE tag with RFID antennas attached on is devised as smart tag to have ability of broadcasting signals while can be sensed by RFID readers simultaneously. Two modes of signal source served as input for the verification of MMBL. Total 190 smart tags are temporarily placed at the centre of each hexagon cell to broadcast signals for calibration purposes, and we retain 6 of tags as the permanent deployed reference tags for online prediction. Both RFID readers and STGs are able to upload scanned data to form datasets in the dimension of $\mathbb{R}^{m\times 15}$ for offline and online stage where m is the numbers of pieces of data. Trolleys and trays are outfitted with smart tag for online prediction purpose.

6.2. Result analysis

The model was developed with the same parameter settings as the laboratory-based model but was adapted to handle different input dimensions, incorporating a total of 15 dimensions of collected data. Fig. 10 employs a cumulative distribution function to compare the location error across different methods. As shown in Fig. 10, MMBL significantly outperforms other methods, with 95% of the location errors contained within 3.41 m, which is more accurate than CNN (5.67m), LSTM (3.53m), and other gradient boosting methods. Additionally, 99% of the location errors for MMBL are within 4.70 m in industrial settings. These results suggest that MMBL possesses strong anti-interference capabilities when dealing with unstable RF environments. In the hexagon cell mapping process, the MMBL achieves an accuracy rate of over 90% for correct classification within its parent and adjacent levels when the inner diameter is greater than half of the 99% location error margin.

Table 3 provides a summary of the positioning results from different

Table 2
Overall positioning results (in meters) in laboratory experiment.

Method	Instant					Rearranged layout with partition panels				
	ME (m)	SD (m)	P_{75}	P_{95}	P_{99}	ME (m)	SD (m)	P_{75}	P_{95}	P_{99}
MMBL	**0.44**	**0.34**	**0.49**	**1.44**	**3.05**	**0.76**	**1.10**	**0.98**	**1.92**	**3.79**
CNN[58]	1.98	1.22	2.70	3.91	4.88	2.62	1.98	3.41	4.66	6.04
LSTM[13]	1.07	0.99	1.32	3.06	5.25	1.68	1.23	1.98	3.78	5.98
CatBoost[59]	2.70	3.45	4.05	5.95	6.68	3.21	4.81	4.74	6.82	7.31
AdaBoost[60]	1.96	1.35	2.41	4.02	4.51	2.49	2.18	3.21	4.61	5.10

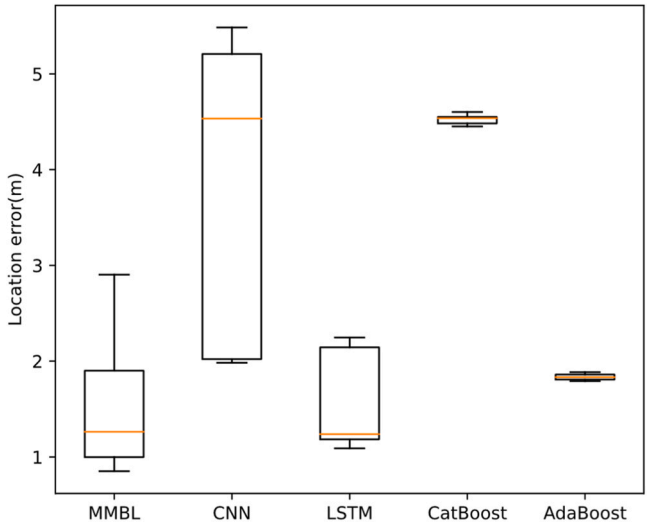

Fig. 8. Boxplot of 10-fold cross-validation for different methods in laboratory experiment.

methods immediately after project implementation and following one year of operation. Initially, MMBL achieves the lowest mean error (ME) of 1.87 m, compared to CNN (2.66m), LSTM (2.14m), CatBoost (4.16m), and AdaBoost (3.31m). Remarkably, after one year, MMBL still displays impressive accuracy, with a mean error of 2.12 m, which is an improvement of over 40% relative to the other methods. The performance of MMBL only degrades by 0.25 m, showcasing the best robustness among the baseline methods. Although LSTM initially exhibits the closest accuracy to MMBL, it suffers a significant increase in mean error to 3.62 m without small-scale re-learning before each prediction. Furthermore, the standard deviation (SD) for MMBL is 1.20 m, lower

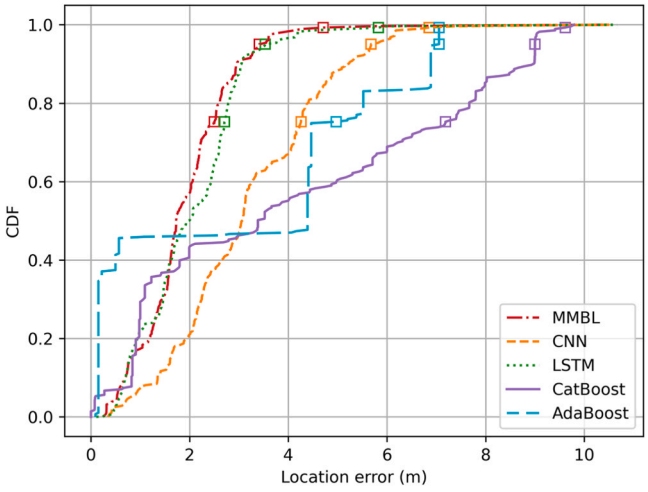

Fig. 10. Comparison of cumulative distribution function of location error for different methods in assembly line area.

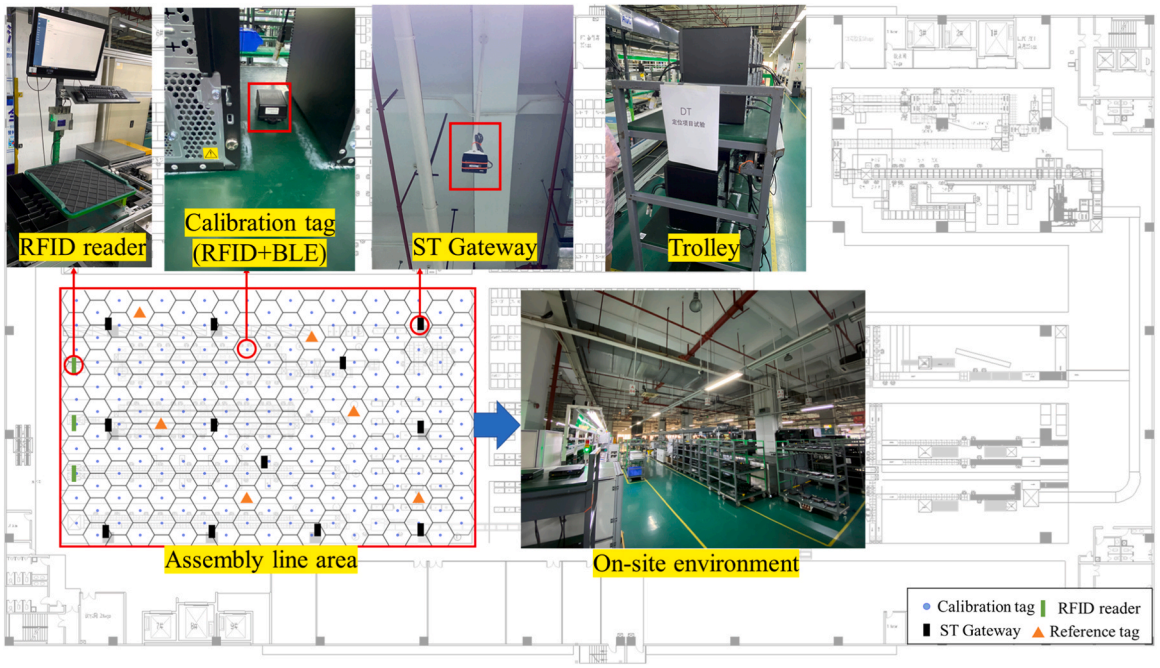

Fig. 9. Factory floor plan and deployment explanation.

Table 3
Overall positioning results (in meters) in assembly line area.

Method	Instant					After one year operation				
	ME (m)	SD (m)	P_{75}	P_{95}	P_{99}	ME (m)	SD (m)	P_{75}	P_{95}	P_{99}
MMBL	**1.87**	**1.20**	**2.49**	**3.41**	**4.70**	**2.12**	**1.85**	**2.68**	**3.68**	**5.23**
CNN[58]	2.66	1.82	4.25	5.67	6.85	4.05	2.92	6.07	9.51	9.95
LSTM[13]	2.14	1.27	2.70	3.53	5.82	3.62	2.12	3.45	5.81	7.64
CatBoost[59]	4.16	6.46	7.18	8.99	9.31	6.92	6.65	7.62	9.45	9.97
AdaBoost[60]	3.31	3.28	4.96	7.05	7.05	4.18	4.82	5.32	7.89	7.89

than that of the other machine learning methods, indicating a more consistent location estimation. Detailed analyses of training efforts and model robustness for the three artificial neural network methods are presented in Fig. 11 and Fig. 12.

Fig. 11 illustrates the convergence of the loss function during the training and validation phases for various ANN methods. By incorporating an early stopping mechanism, which halts training when performance on the validation dataset begins to decline, overfitting is mitigated. The MMBL model demonstrates rapid convergence, requiring approximately 260 epochs to reach the point of minimum loss. This is significantly more efficient than both CNN and LSTM, which converge at 965 epochs and 1000 epochs, respectively.

Fig. 12 presents the results of 10-fold cross-validation for different methods in assembly line area using boxplot. We conduct 10 training iterations, each with a different subset of the training data, to evaluate the model's performance stability. Under the industrial settings, the median value of MMBL is significantly lower than other methods. The LSTM, CatBoost, and AdaBoost are more than other models but exhibit higher location errors that MMBL. The proposed MMBL model shows low variance across the multiple training trials and consistently achieves lower location estimation errors.

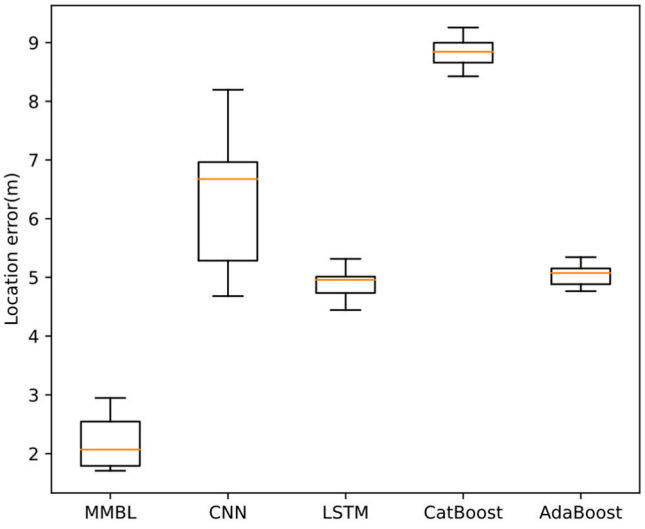

Fig. 12. Boxplot of 10-fold cross-validation for different methods in assembly line area.

7. Conclusion

This research first investigates the sharable and interoperable spatial-temporal elements in the cyber-physical internet inspired by the computer networks and computer architecture. Spatial-temporal point of services is elaborated and modelled to specify contextual events and transactions of industrial resources for spatial-temporal analytics. The MMBL model incorporating multiple existing sensors from CPIS is developed to estimate the location of objects in industrial settings as analogous to biological cells mutation and evolution. Reference tags with small-scale re-learning are developed for durable and reliable positioning. Three kinds of spatial-temporal reasoning are proposed to identify the pattern and rules through the spatial-temporal information of the objects which provides references for decision-making process. We implement the solution and benchmark MMBL against other machine learning based indoor positioning algorithms in both laboratory experiment and a real-life case study. Results shows that the MMBL has better performances than other models in terms of instant and long-term accuracy, convergence speed and robustness, which supports the spatial-temporal traceability for CPIS.

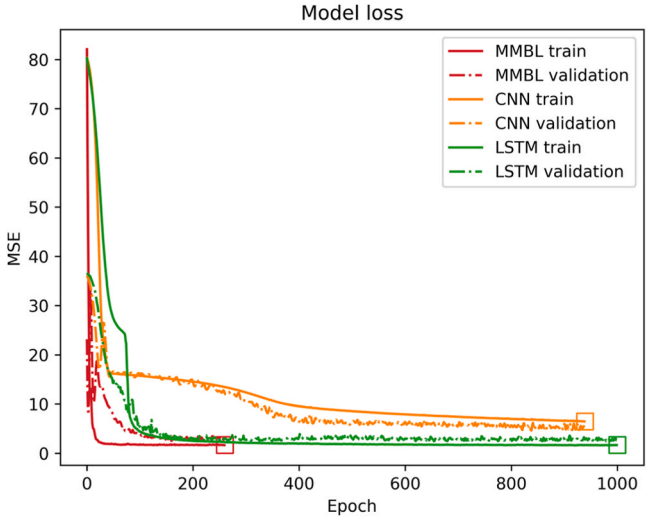

Fig. 11. Convergence curves of three models on training and validation data sets.

The novelty of this research is threefold. Firstly, the proposed unified spatial-temporal elements link and standardize the indoor and outdoor spatial representations and temporal measurement for better interoperability and sharing across the global supply chain. Secondly, inspired by the mutation and evolution of bionic cells, the MMBL innovatively blends multi-modal signals from CPIS into a machine learning model to ensure accuracy and enduring effectiveness of indoor positioning. Thirdly, the spatial-temporal reasoning mechanisms contribute to mitigate the effects of uncertainties and disruptions that occur during operations. Insights extracted from seemingly unrelated spatial-temporal data support the resilience of real-time decision-making. These spatial-temporal reasoning mechanisms serve as a bridge, transitioning us from the technology-centric focus of Industry 4.0 to the value-driven perspective of Industry 5.0.

Finally, several opportunities for future research can be identified as follows: Standard representation of spatial-temporal data provides unified interface for message sharing among encrypted method such as blockchain. Future research should come up with encrypted technology to handle the tempering, mutability, and repudiation issues for spatial-temporal traceability sharing purpose. The calibration of indoor positioning is labour-intensive and time-consuming, future research could be directed towards identifying the intrinsic connections to lower the calibration efforts. With the unprecedented large-scale application of

industrial IoT, the real-time status of resources can provide online adjustment for dynamic scheduling as the spatial-temporal related uncertainties and disturbances can be confirmed and mitigated to some degrees. The performance of dynamic scheduling can be further enhanced by considering spatial-temporal driven knowledge.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported by Natural Science Foundation of China (no. 52305557), Start-up Fund of Hong Kong Polytechnic University (no. P0046067), Open Fund of State Key Laboratory of Intelligent Manufacturing Equipment and Technology (no. IMETKF2024022), Departmental General Research Fund of the Hong Kong Polytechnic University (no. P0045749), China Postdoctoral Science Foundation (no. 2022M712394, 2023M730406), Hong Kong RGC TRS Project (T32-707/22-N), Research Impact Fund (R7036-22), and Collaborative Research Fund (C7076-22GF).

References

[1] Montreuil B. Toward a Physical Internet: meeting the global logistics sustainability grand challenge. Logist Res 2011;3(2):71–87.
[2] Pan S, Zhong RY, Qu T. Smart product-service systems in interoperable logistics: design and implementation prospects. Adv Eng Inform 2019;42:100996.
[3] Mervis J. The information highway gets physical. Science 2014;344(6188):1104–7.
[4] Ivanov D, Dolgui A. Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak. Int J Prod Res 2020;58(10):2904–15.
[5] Ballot E, Montreuil B, Zacharia ZG. Physical Internet: First results and next challenges. J Bus Logist 2021;42(1):101–7.
[6] Christopher M. Logistics & Supply Chain Management. Uk: Pearson; 2016.
[7] Langley CJ, et al. Supply Chain Management: A Logistics Perspective. Cengage Learning; 2020.
[8] Grünwald HJ, Fortuin L. Many steps towards zero inventory. Eur J Oper Res 1992;59(3):359–69.
[9] Lyu Z, et al. Towards Zero-Warehousing Smart Manufacturing from Zero-Inventory Just-In-Time production. Robot Comput-Integr Manuf 2020;64:101932.
[10] Zumberge J, et al. Precise point positioning for the efficient and robust analysis of GPS data from large networks. J Geophys Res: Solid earth 1997;102(B3):5005–17.
[11] Chen Y, et al. The impact of gis/gps network information systems on the logistics distribution cost of tobacco enterprises. Transp Res Part E: Logist Transp Rev 2021;149:102299.
[12] Zhou M, et al. Integrated Statistical Test of Signal Distributions and Access Point Contributions for Wi-Fi Indoor Localization. IEEE Trans Veh Technol 2021;70(5):5057–70.
[13] Hoang MT, et al. Recurrent neural networks for accurate RSSI indoor localization. IEEE Internet Things J 2019;6(6):10639–51.
[14] Daránvi A, et al. Processing indoor positioning data by goal-oriented supervised fuzzy clustering for tool management. J Manuf Syst 2022;63:15–22.
[15] Feng D, et al. Kalman filter based integration of IMU and UWB for high-accuracy indoor positioning and navigation. IEEE Internet Things J 2020. p. 1-1.
[16] Zhang Y, Qu C, Wang Y. An indoor positioning method based on CSI by using features optimization mechanism with LSTM. IEEE Sens J 2020;20(9):4868–78.
[17] Chen Z, et al. A novel real-time deep learning approach for indoor localization based on rf environment identification. IEEE Sens Lett 2020;4(6):1–4.
[18] Wang P, Morton YJ. Multipath estimating delay lock loop for LTE signal TOA estimation in indoor and urban environments. IEEE Trans Wirel Commun 2020;19(8):5518–30.
[19] Xi Y, et al. Beyond the First Law of Geography: Learning Representations of Satellite Imagery by Leveraging Point-of-Interests. *Proc ACM Web Conf 2022* 2022.
[20] Yuan, Q., et al., *Time-aware point-of-interest recommendation*, in Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. 2013, Association for Computing Machinery: Dublin, Ireland. p. 363–372.
[21] Molina B, et al. A multimodal fingerprint-based indoor positioning system for airports. IEEE Access 2018;6:10092–106.
[22] Kuo RJ, et al. The application of an artificial immune system-based back-propagation neural network with feature selection to an RFID positioning system. Robot Comput-Integr Manuf 2013;29(6):431–8.
[23] Lu S, et al. A passive RFID tag-based locating and navigating approach for automated guided vehicle. Comput Ind Eng 2018;125:628–36.
[24] Zhao Z, et al. Logistics sustainability practices: an IoT-enabled smart indoor parking system for industrial hazardous chemical vehicles. Int J Prod Res 2020:1–17.
[25] Wu W, et al. Industrial IoT and long short-term memory network enabled genetic indoor tracking for factory logistics. IEEE Trans Ind Inform 2022. p. 1-1.
[26] Arslan M, Cruz C, Ginhac D. Semantic enrichment of spatio-temporal trajectories for worker safety on construction sites. Pers Ubiquitous Comput 2019;23(5):749–64.
[27] Lingyun, Y., et al. *RFID data fusion algorithm based on spatio-temporal semantics in internet of things*. in 2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI). 2017.
[28] Jin, B. and H. Chen. *Spatio-Temporal Events in the Internet of Things*. in *2010 IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*. 2010.
[29] De, S., et al. *Service modelling for the Internet of Things*. in *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*. 2011.
[30] Bernardo R, Sousa JMC, Gonçalves PJS. Survey on robotic systems for internal logistics. J Manuf Syst 2022;65:339–50.
[31] Zhong Y, et al. Image-based flight control of unmanned aerial vehicles (UAVs) for material handling in custom manufacturing. J Manuf Syst 2020;56:615–21.
[32] Perez-Grau FJ, et al. Introducing autonomous aerial robots in industrial manufacturing. J Manuf Syst 2021;60:312–24.
[33] Danys L, et al. Visible Light Communication and localization: a study on tracking solutions for Industry 4.0 and the Operator 4.0. J Manuf Syst 2022;64:535–45.
[34] Popović G, et al. Human localization in robotized warehouses based on stereo odometry and ground-marker fusion. Robot Comput-Integr Manuf 2022;73.
[35] Gholami M, Cai N, Brennan RW. An artificial neural network approach to the problem of wireless sensors network localization. Robot Comput-Integr Manuf 2013;29(1):96–109.
[36] Bougdira A, Akharraz I, Ahaitouf A. A traceability proposal for industry 4.0. J Ambient Intell Humaniz Comput 2020;11(8):3355–69.
[37] Schabus S, Scholz J. Geographic Information Science and technology as key approach to unveil the potential of Industry 4.0: How location and time can support smart manufacturing. 2015 12th Int Conf Inform Control, Autom Robot (ICINCO) 2015.
[38] Kumar, S.A.P. and M.A. Brown. *Spatio-Temporal Reasoning within a Neural Network framework for Intelligent Physical Systems*. in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2018.
[39] Jayanthi G, Uma V. Event attributed Spatial Entity Knowledge (EASE) based Spatio-Temporal reasoning to infer geographic processes. *2016 2nd Int Conf Green High Perform Comput (ICGHPC)* 2016.
[40] Wang F, Huang QY. The importance of spatial-temporal issues for case-based reasoning in disaster management. *2010 18th Int Conf Geoinformatics* 2010.
[41] Han, J., M. Kamber, and J. Pei, *13 - Data Mining Trends and Research Frontiers*, in *Data Mining (Third Edition)*, J. Han, M. Kamber, and J. Pei, Editors. 2012, Morgan Kaufmann: Boston. p. 585–631.
[42] Li M, et al. Spatial–temporal out-of-order execution for advanced planning and scheduling in cyber-physical factories. J Intell Manuf 2022;33(5):1355–72.
[43] Zhai Y, et al. Spatial-temporal hedging coordination in prefabricated housing production. Int J Prod Econ 2020;229:107792.
[44] Zhao Z, et al. Cyber-physical spatial temporal analytics for digital twin-enabled smart contact tracing. Ind Manag Data Syst 2021;121(5):1082–106.
[45] Zhao Z, et al. Digital twin-enabled dynamic spatial-temporal knowledge graph for production logistics resource allocation. Comput Ind Eng 2022;171:108454.
[46] Iqbal R, et al. Fault detection and isolation in industrial processes using deep learning approaches. IEEE Trans Ind Inform 2019;15(5):3077–84.
[47] Bondaruk, B., S. Roberts, and C. Robertson, *Discrete global grid systems: Operational capability of the current state of the art*. Spatial knowledge and information Canada, 2019.
[48] Wang X, et al. A storage method for remote sensing images based on google S2. IEEE Access 2020;8:74943–56.
[49] Sahr K. Central place indexing: Hierarchical linear indexing systems for mixed-aperture hexagonal discrete global grid systems. Cartogr: Int J Geogr Inf Geovisualization 2019;54(1):16–29.
[50] Uber. S2 | H3. 2022; Available from: ⟨https://h3geo.org/docs/comparisons/s2⟩.
[51] Zhong RY, et al. Visualization of RFID-enabled shopfloor logistics big data in cloud manufacturing. Int J Adv Manuf Technol 2016;84(1):5–16.
[52] Yang C, Shen W, Wang X. The internet of things in manufacturing: key issues and potential applications. IEEE Syst, Man, Cybern Mag 2018;4(1):6–15.
[53] Sánchez P, Bellogín A. Point-of-interest recommender systems based on location-based social networks: a survey from an experimental perspective. ACM Comput Surv 2022.
[54] Tao F, et al. Digital twin and its potential application exploration. Comput Integr Manuf Syst 2018;24(1):1–18.
[55] Gyulai D, Pfeiffer A, Bergmann J. Analysis of asset location data to support decisions in production management and control. Procedia CIRP 2020;88:197–202.
[56] Abdirad M, Krishnan K. Industry 4.0 in logistics and supply chain management: a systematic literature review. Eng Manag J 2021;33(3):187–201.
[57] Shen Z, et al. When RSSI encounters deep learning: an area localization scheme for pervasive sensing systems. J Netw Comput Appl 2021;173.
[58] Ibrahim M, Torki M, ElNainay M. CNN based indoor localization using RSS time-series. *2018 IEEE Symp Comput Commun (ISCC)* 2018.
[59] Prokhorenkova L, et al. CatBoost: unbiased boosting with categorical features. Adv Neural Inf Process Syst 2018:**31**.
[60] Hastie T, et al. Multi-class adaboost. Stat its Interface 2009;2(3):349–60.